

Stefan Gradmann

B 7 Semantic Web und Linked Open Data

B 7.1 Was ist das „Semantic Web“?

Obwohl häufig als „Web 3.0“ apostrophiert ist das „Semantic Web“ kein Alternativentwurf zum ursprünglichen World Wide Web (WWW) und auch keine neue Version desselben, die in eine auf dem Social Web (vulgo Web 2.0) aufbauende Kontinuität eingeschrieben wäre, sondern das Semantic Web ist am ehesten als Erweiterung und Fortentwicklung des Web der ersten Generation zu verstehen. In den Worten des „Erfinders“ sowohl des WWW als auch des „Semantic Web“: „The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.“ (Lit. 07) Für das Verständnis des Semantic Web essentiell ist mithin ein Grundverständnis der Informationsarchitektur des WWW der ersten Generation und seiner Erweiterungen im Semantic Web, gleichermaßen der zweiten Generation des WWW. Ohne dass diese hier im Detail referenziert würden, verdanken sich Teile der nachstehenden Darstellung drei grundlegenden, lehrbuchartigen Einführungen: Antoniou & Van Harmelen (Lit. 01), Hitzler et al. (Lit. 16) und Heath, Bizer (Lit. 14). Alle drei sind auch für eine vertiefende Einführung empfehlenswert.

B 7.1.1 Das Web der ersten Generation als Web der Dokumente

Die in Abbildung 1 sichtbare Struktur ist diejenige des WWW der ersten Generation: in ihrem Zentrum steht „This document“. Es handelt sich um Tim Berners-Lees „Proposal“ (Lit. 04) an seinen Arbeitgeber CERN mit dem Ziel, eine Plattform für kollaboratives Informationsmanagement aufzubauen, in gewisser Hinsicht die Geburtsurkunde des WWW der ersten Generation, des „Document Web“. Dies war konzipiert als eine gigantische Hypertextanwendung mit für Menschen lesbaren Dokumenten und Verbindungen zwischen diesen.

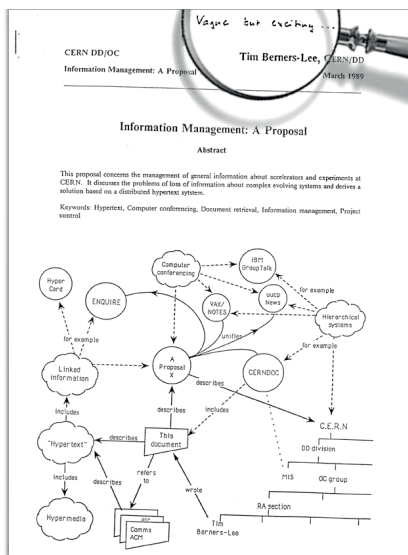


Abb. 1: Information Management: A Proposal (Dokument von Tim BL mit handschriftlicher Notiz von Mike Sendall)

Das änderte sich dann mit der ersten Implementierung des WWW, der eine Architektur zugrunde liegt, deren Kern durch den Graphen in Abbildung 2 illustriert ist:

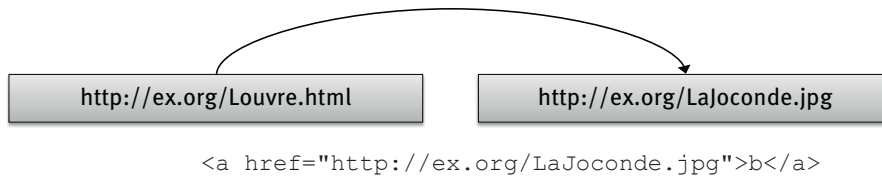


Abb. 2: Ein HTTP-Graph

Zu sehen sind zwei WWW-Ressourcen mit URIs aus einem imaginären Namensraum `ex.org`, die durch einen HTTP-Link verbunden sind. Dies ist ein typisches Beispiel des Web der (menschenslesbaren) Dokumente, bei dem wir mit unserem kulturellen Hintergrund natürlich begründete Vermutungen darüber anstellen können, was für Arten von Entitäten hier referenziert sind: wir wissen, dass „Louvre“ eine Instanz der Klasse „Museum“ ist, wir wissen zumeist auch, dass „La Joconde“ die in Frankreich gängige Bezeichnung für das Bild ist, das wir im deutschen Sprachraum kurz als „Mona Lisa“ bezeichnen, dass es sich also um eine Instanz der Klasse „Gemälde“ handelt. Eine Maschine jedoch ist völlig außerstande, Vermutungen dieser Art anzustellen.

Und weiter: Wir wissen natürlich, dass Gemälde in Museen hängen, können also aus unserem allgemeinen Wissen über die Beziehungen zwischen Instanzen der Klassen „Gemälde“ und „Museen“ Schlüsse ableiten hinsichtlich der wahrscheinlichen Beziehung zwischen den hier referenzierten beiden konkreten Instanzen: Das Bild „La Joconde“ hängt mit großer Wahrscheinlichkeit im Louvre. Auch Schlüsse dieser Art kann eine Maschine nicht ziehen – es sei denn, deren Voraussetzungen würden ihr explizit mitgeteilt. Dafür jedoch fehlt es dem WWW der ersten Generation schlicht an Ausdruckskraft: Weder kann dort die Beziehung zwischen einer Klasse und ihren Instanzen zum Ausdruck gebracht werden, noch lassen sich Beziehungen zwischen Web-Ressourcen typisieren. Dies ermöglicht erst eine syntaktische Erweiterung der WWW-Architektur, die zugleich der Grundbaustein des „Semantic Web“ ist: das *Resource Description Framework* (RDF) und seine „Grammatik“, das *RDF Schema* (RDFS).

B 7.1.2 RDF und RDFS als syntaktische Erweiterung des WWW

In RDF kann die Tatsache, dass La Joconde im Museum „Musée du Louvre“ hängt, durch die folgende Aussage explizit gemacht werden:



Abb.3: Ein RDF-Tripel

Und auch die Zuordnung dieser Instanzen zu Klassen wird in RDF formulierbar:

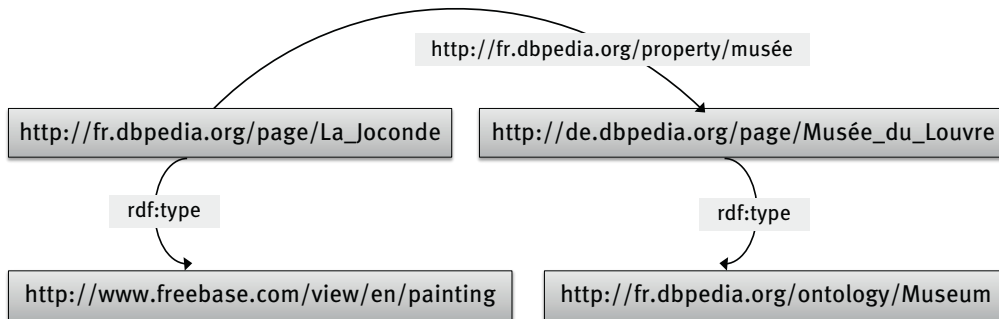


Abb. 4: Instanzen und Klassen in RDF

Und schließlich kann in RDF natürlich auch die allgemeine Beziehung zwischen den Instanzen zweier Klassen ausgedrückt werden:

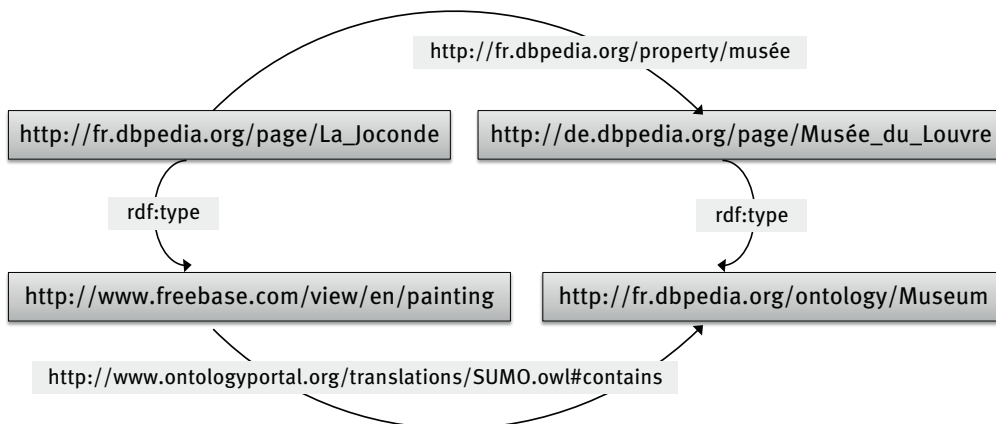


Abb. 5: Beziehung zwischen Instanzen zweier Klassen in RDF

Und aus dieser letzten Aggregation von RDF-Aussagen hätte die erste (<La_Joconde> - <musée> - <Musée_du_Louvre>) auch von einer Maschine per Inferenz abgeleitet werden können, sie ist also in diesem Aussagenkontext gar nicht mehr explizit erforderlich.

Dafür sorgt die „Grammatik“ RDFS, welche Mengen von RDF-Aussagen in einem Aussagensystem organisierbar macht. Die RDF-Aussagen selbst sind dabei einfache Sätze mit der immer gleichen Struktur Subjekt – Prädikat – Objekt (oder auch Resource – Property – Value), deren erste beiden Elemente Entitäten im WWW sein müssen (und mithin eine URI haben), während an der dritten Position eine WWW-Resource oder eine Zeichenkette (ein so genannter „Literal“) stehen kann. Die Bestandteile solcher auch als „Tripel“ bezeichneten Aussagen werden in RDFS in einem Klassenkonzept organisierbar (<La_Joconde> - <ist vom rdf:type> - <painting>), gleiches gilt für Eigenschaften (<Lehren> - <subPropertyOf> - <Kommunizieren>), und in beiden Fällen sorgt ein einfaches Vererbungskonzept dafür, dass aus solchen Hierarchien von Klassen und Eigenschaften einfache, deterministische Schlussfolgerungen wie im obigen Beispiel gezogen werden können.

B 7.1.3 Die Erweiterung des Repräsentationsraums des WWW: das Web der Dinge

Tatsächlich ermöglicht wird allerdings ein Teil der Aussagen im obigen Beispiel erst durch eine zweite Erweiterung des WWW, diesmal in dessen Geltungsbereich. Die Entitäten des WWW der ersten Generation waren „Dokumente“, im Verständnis des für die Standardisierung im WWW zuständigen World Wide Web Consortium (W3C, <http://www.w3.org/>), Informationsentitäten also, die genuiner Bestandteil dieser gigantischen Hypertextanwendung waren. Das so genannte „Web der Dinge“ nun erweitert den Geltungsbereich des WWW auf die gesamte Welt: Alle Bestandteile unserer Wirklichkeit sollen nun Teil des WWW werden – oder dort zumindest repräsentiert werden können und damit überhaupt erst zum Gegenstand von Aussagensystemen in RDF werden – wie in unserem Beispiel die Resource http://fr.dbpedia.org/page/Musée_du_Louvre, welche zum Beispiel eine `<prop-fr:latitude>` mit dem Wert „48.861073 (xsd:double)“ hat (Beispiel für eine typisierten Literal) oder eine Eigenschaft `<dbpedia-owl:country>` mit dem Wert `<dbpedia-fr:France>`.

B 7.1.4 Das Semantic Web

Mit diesen beiden Erweiterungen des WWW der ersten Generation wird das Postulat (Lit. 07) sehr viel besser einzuordnen: Es geht um nicht weniger als um die Transformation der Hypertext-Anwendung WWW in einen weltumspannenden Wissensraum, in dem die Repräsentation und Kontextualisierung sowohl genuiner WWW-Ressourcen als auch beliebiger Wirklichkeitsausschnitte möglich wird. Die Beziehungen zwischen den dort repräsentierten Entitäten und auch deren Zuordnung zu einem hierarchischen Klassenmodell sind unter Zuhilfenahme so genannter Ontologien (vgl. B 6 Ontologien) so typisierbar, dass eine maschinelle Weiterverarbeitung dieses Beziehungsgeflechts möglich wird und nicht mehr bloß deren intellektuelle Rezeption durch den Menschen – und diese Eigenschaft ist es, die Berners-Lee als „semantisch“ bezeichnet hat. Er hat damit ein weit reichendes Missverständnis begründet, denn mit „Bedeutung“ und „Interpretation“ im landläufigen Sinne haben diese typisierten Beziehungen und deren maschinelle Weiterverarbeitung im Kern wenig zu tun!

Besonders virulent wurde dies Missverständnis auch dadurch, dass das Semantic Web vor allem in seiner Frühzeit an einer Überdominanz des Paradigmas der Künstlichen Intelligenzforschung (KI) beinahe zugrunde gegangen wäre. Dies zumindest ist die von Harmelen (Lit. 13) vertretene Auffassung, der auch der Verfasser dieses Beitrages in seiner Antrittsvorlesung (Lit. 11) beipflichtet. Stärker auf eine Zusammenarbeit mit der KI ausgerichtet sind dem gegenüber Hendler und Berners-Lee (Lit. 15). Fast verwirrend schließlich ist Berners-Lee (Lit. 05), wo er unter dem Titel „Artificial Intelligence and the Semantic Web“ einleitend feststellt, dass „SW is not AI and AI is not SW“, um dann festzuhalten, welches Erbe das Semantic Web von der KI übernommen hat und schließlich auf der mit „Distractions: Meaning of meaning“ überschriebenen Seite 20 mit unklarer Intention und ohne erkennbares Ergebnis über „Bedeutung“ zu rasonieren.

Fakt ist jedenfalls, dass die auf komplexen Logikansätzen basierenden Versprechungen der Frühzeit des Semantic Web mit ihren intelligenten Agenten und einer großen Nähe zu den weit überzogenen Versprechen der KI das Semantic Web große und kaum einlösbare Erwartungen hinsichtlich der Repräsentierbarkeit von Bedeutung und Interpretation im menschlichen Sinne erzeugt haben, die der erfolgreichen Entwicklung des Projektes nicht eben förderlich waren.

Dem gegenüber ist der proklamierte Ansatz der Linked Open Data (Lit. 02, s. unten) ein deutlich bescheidenerer Ansatz der Wissensgenerierung durch Kontextualisierung (vgl. dazu Lit. 12), der aufgrund seiner deutlich geringeren Komplexität und größeren Skalierbarkeit zumindest für eine Übergangszeit eine klar größere Breitenwirkung erzeugt hat, als dies das „Semantic Web“ je vermocht hatte.

B 7.2 Das Schichtenmodell des Semantic Web

Schon in der ersten, kanonischen Darstellung (Lit. 07) findet sich als Ansatz für die systematische Beschreibung des dort postulierten Ansatzes ein Schichtenmodell, das für die Darstellungstradition des Semantic Web so weitgehend paradigmengestaltend geworden ist, dass es dann gar (Lit. 15) parodiert werden konnte. Abbildung 6 zeigt dieses erste Schichtenmodell:

Dies erste Modell baut noch deutlich erkennbar auf der Dokumentmetapher auf, indem die beiden untersten, tragenden Schichten der Codierung (in Unicode) und Referenzierung von Entitäten (mit URIs) sowie der damals noch exklusiven Formulierungssyntax (mit XML, Namensräumen [NS] und XML-Schema) die Basis legen für etwas, was Berners-Lee et al. als „Self Describing Document“ bezeichnen. RDF und RDF-Schema bauen auf diesen Elementarschichten und auf RDF wiederum – nun auf der Ebene von „Daten“ – die in Ontologien des Semantic Web organisierte Begrifflichkeit, mit deren Hilfe – nunmehr auf der Ebene von „Regeln“ – logische Operationen möglich werden. Die Spitze dieses ersten Schichtenmodells bilden – damals eigentlich nur als Forschungsbereiche benannt – „Beweisführung“ und „Vertrauen“. Digitale Unterschriften durchziehen als vertikale Authentifizierungstechnologie die zentralen Schichten dieser „Torte“, ihr Geltungsbereich endet jedoch rätselhafterweise direkt unterhalb von „Vertrauen“.

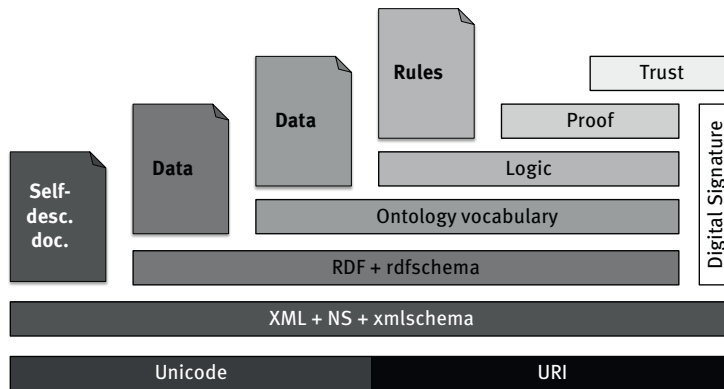


Abb. 6: Das erste Schichtenmodell des Semantic Web (© Berners-Lee et al. 2001, Lit. 07)

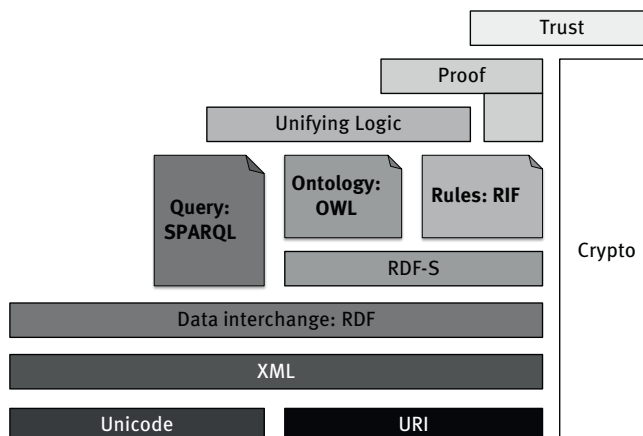


Abb. 7: User interface & applications: das Schichtenmodell des W3C

Eine zweite Fassung dieses Schichtenmodells aus dem Jahr 2006 – nunmehr in einer „kanonischen“, vom W3C verabschiedeten Version – verschiebt gegenüber der ersten die Gewichte und führt neue Schichten ein, wie Abbildung 7 zeigt.

Unicode ist aus der „Trägerschicht“ verschwunden, allein noch die WWW-spezifische URI ist hier angesiedelt. Darüber liegt zwar immer noch XML – nun aber nur noch als eine Vermittlungsoption zu RDF: Andere, hier nicht explizit benannte Syntaxvarianten wie N3 oder Turtle, ermöglichen ggf. einen alternativen Zugang zu RDF als der eigentlichen Trägerschicht, oberhalb der das Bild nun deutlich differenzierter wird. Neben die wissensmodellierenden Schichten des ursprünglichen Modells mit RDFS und der 2001 noch gar nicht existierenden *Ontology Working Language* (OWL) tritt nun als Standard-Abfragesprache SPARQL und als vermittelnde Alternative zu den weitgehend unveränderten oberen Schichten das *Rule Interchange Format* (RIF). Vor allem aber ist die neu hinzugekommene Schicht „*User Interface & Applications*“ deutliches Zeichen für die nun beginnende Anwendungsreife der semantischen Technologien.

Dem gegenüber deutlich komplexer schließlich ist das würfelartig angeordnete Modell von Nowack (Lit. 17), in dem Spezifikationen und (technische) Lösungen aufgeführt sind, während Konzepte und Abstraktionen zu Funktionen wie Abfrage und Applikationen vermitteln.

Dabei haben sich insbesondere die Serialisierungsformate von RDF deutlich vermehrt: Hinzugekommen sind Turtle, RDFa und MicroFormats. Und auch die zentrale Rolle von RDF selbst ist durch seine prominente Stellung im Bereich der „Concepts & Abstractions“ nochmals deutlich betont.

B 7.3 Linked Open Data

„I heard someone say that ‘Linked Data’ was an explicit attempt to rebrand semweb and RDF, so you’re right, no difference.“ (O’Reilly, Lit. 18)

Im Modell von Nowack steht hierzu nur lapidar vermerkt: „Linked Data uses a small selection of technologies.“ Und auch die diesem Abschnitt als Motto vorangestellte Äußerung von O’Reilly lässt vermuten, dass es sich bei Linked Open Data eigentlich nur um eine Art „Semantic Web Light“ handelt. Doch so ganz stimmt das nicht: So sagt etwa Berners Lee (Lit. 06), Linked Data sei „Semantic Web done right“ – eine interessante Äußerung, die insbesondere offen lässt, was denn dann vorher offenkundig nicht richtig gemacht worden ist. Neben der schon mehrfach angesprochenen Überfrachtung des Semantic Web durch schwergewichtige, logikbasierte Ansätze der KI dürfte der Hintergrund wohl vor allem sein, dass Berners-Lee seit ungefähr 2006 den Begriff „Semantic Web“ praktisch nicht mehr verwendet, und in seinem Blog-Eintrag vom 21.11.2007 auch so etwas wie eine Begründung dafür liefert:

„I called this graph the Semantic Web, but maybe it should have been Giant Global Graph! Any worse than WWW? ;-) Not the „Semantic Web“ term has been established for a long time, I’m not proposing to change it. But let’s think about the graph which it is.“ (Lit. 03)

„Linked Open Data“ kann also durchaus auch als der Versuch gesehen werden, das mit Gebrauch des Attributs „semantic“ in die Welt gesetzte Missverständnis wieder auszuräumen.

Vor allem aber ist es der Versuch, das Projekt des Semantic Web, das damals wie in Lit. 13 geschildert eher flügelahm war, nun durch eine wirklich signifikante, frei verfügbare Datenbasis wirklich effektiv zu machen, und dies geschah vor allem durch eine radikale Vereinfachung der zugrunde liegenden Funktionsprinzipien in Lit. 05 von Berners-Lee. Letztlich basiert der gesamte Ansatz der Linked Open Data auf den folgenden vier hier knapp erläuterten Prinzipien:

1. Use URIs as names for things.

Dieser Grundsatz basiert auf der oben erwähnten Erweiterung im Geltungsbereich des WWW: Alles (so genannte *information resources* aber auch wirkliche Dinge in der Welt, die *non infor-*

mation resources) soll nun im WWW repräsentierbar und durch eine URI identifizierbar werden. Ein inzwischen weitgehend verbreiteter Grundsatz.

2. Use HTTP URIs so that people can look up those names
Sind Dinge einmal mit URIs identifiziert empfiehlt der zweite Grundsatz, sie mit Standard-Zeichern des HPPP-Protokolls zu referenzieren – auch dies ein weitgehend durchgesetzter und vergleichsweise trivialer Grundsatz.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
Standards wie SPARQL und RDF sollen letztlich eine API auf Datenebene vorgeben, über die unabhängig von partikularen Applikationen auf die Daten zugegriffen werden kann.
4. Include links to other URIs, so that they can discover more things.
Dieser letzte Grundsatz macht aus offenen Daten im Web Linked Data: die Schaffung von möglichst vielen Verbindungen zu anderen Entitäten im WWW, deren maximale Kontextualisierung im Giant Global Graph.

Die konsequente Berücksichtigung dieser vier einfachen Grundsätze macht tatsächlich aus dem WWW einen „Global Data Space“ (so der Titel von Heath & Bizer, Lit. 14).

B 7.3.1 Offenheit

Offen blieb lange Zeit der Anspruch des Attributs „offen“. In der oben schon zitierten Präsentation (Lit. 06) etwa sagt Berners-Lee einerseits „Free is good“ – andererseits aber auch „For-pay is ok, but so much less enabling“, um dann auf Folie 22 ähnlich ambivalent abzuschließen mit den beiden Sätzen „You must use the open standards RDF, SPARQL“ und „Openness (allowing access) is separate question.“

2010 hat Berners-Lee dann seinen Ansatz auf der Design-Issues-Seite nochmals erweitert, um diese Frage der „Offenheit“ weiter zu klären – er macht in der unten stehenden Tabelle zwar eine offene Lizenzform zur Voraussetzung für den ersten und elementaren Stern, erweitert dann aber in der Folge technische Offenheit zum Schwerpunkt: die Reihung strukturierte Daten → nichtproprietär strukturierte Daten → Referenzierung mittels W3C-Standards → Verlinkung ist auf technische Offenheit, auf transparente Prozessierbarkeit fokussiert.

Sterne	Voraussetzung
★	Available on the web (whatever format) but with an open licence, to be Open Data
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

Tab. 1: Voraussetzungen nach Berners-Lee

Die Frage Verfügbarkeitsbedingungen bleibt damit nur vage beantwortet – und damit eine der offenen Flanken des LoD-Ansatzes, insbesondere im Bereich der Public Sector Information (PSI), also der vom öffentlichen Sektor produzierten Rohdaten (Geodaten, Haushaltsdaten u. a. m.), die zunehmend auf den öffentlichen Markt kommen – und zunehmend eben auch als Linked Data mit dann zu klärenden Verfügbarkeitsbedingungen.

Insbesondere aufgrund der relativ offensiven Haltung der Europäischen Kommission in dieser Frage (s. dazu unter http://ec.europa.eu/information_society/policy/psi/index_en.htm) sind momentan dedizierte Forschungs- und Vernetzungsprojekte wie LAPSI (<http://www.lapsi-project.eu/>) ausschließlich mit rechtlichen Aspekten von PSI befasst.

B 7.3.2 Qualität

Linked Open Data ist ein ganz erheblich von der Europäischen Kommission geförderter Ansatz, und die beiden wahrscheinlich technologisch treibenden Projekte waren europäisch gefördert: das Linked-Open-Data-Projekt hat im großen Maßstab Datenressourcen hervorgebracht und verlinkt – das Resultat ist die allseits bekannte „Wolke“. Und das Projekt LOD2 (<http://lod2.eu/>) hat bezogen auf diese Wolke einen beeindruckenden „Stack“ an Technologien entwickelt.

Je besser nun aber die Übersicht über die nun erreichte Datendichte wird – zu nennen wären hier vor allem Dienste wie [VOV](http://lov.okfn.org/dataset/lov/) (<http://lov.okfn.org/dataset/lov/>) und der *Falcons Ontology Search* unter <http://ws.nju.edu.cn/falcons/ontologysearch/index.jsp>, desto deutlicher wird das mit diesem Ansatz aufgebaute Qualitätsproblem: Die Ontologien und Vokabulare sind weitgehend unabhängig voneinander entstanden und semantisch meist teilredundant. Der einzige Konstruktor, den die Ontologie-Konstruktionssprache OWL für die Behebung dieses Problems hat, ist `owl:sameAs` – und dieser ist in den allerwenigsten Fällen wirklich angemessen.

Eine Suche nach Klassenmodellierungen des Konzeptes „Freedom“ in Falcons etwa fördert 10 verschiedene ontologische Quellen zutage, und die wenigsten Klassen-Instanzen in diesen Quellen können wirklich mit `owl:sameAs` miteinander verbunden werden! In diesem Bereich, dem Ontology-Mapping sowie der Bewertung und Modellierung semantischer Nähe und Überschneidung, liegt momentan eines der aktivsten Forschungsfelder der LoD-Community.

B 7.4 Semantic Web und Anwendungen mit Linked Open Data

Anwendungen auf Basis von Technologien des Semantic Web und der Linked Data sind inzwischen aus dem WWW-basierten Dienstleistungsbereich kaum mehr wegzudenken. Das gilt sowohl für die unzähligen Reise- und Dienstleistungsportale, die meist in einer Kombination mit Techniken des Social Web mehr oder minder ausgeklügelte Vermutungen über die Bedarfe ihrer Nutzer anstellen als auch für Produktkalkulationen und -optimierungen logische Kalküle verwenden, die auf Inferenzen über eine Vielzahl in RDF gefasster Einzelaussagen basieren.

Dass aus der Kombination von Semantic Web und der Linked Data schon ein relativ wissenschaftsnahes Anwendungsbeispiel aus der Industrie entstehen kann, zeigt der Ansatz der Firma Biogen Idec (Lit. 09). Mit Hilfe des kommerziellen, Semantic Web-basierten Produktes Anzo der Fa. Cambridge Semantics konnte die eigene komplexe Liefer- und Vertragskette so optimiert werden, dass subjektiv ein enormer Gewinn an Flexibilität und Effizienz die Folge war – allerdings ohne dass der Sprecher der Firma, Chris Beganski, diesen Gewinn quantifizieren könnte. Hier die Begründung: „How do you put a value on now being able to know something that you didn't know before?“

Vielleicht mit das prominenteste Anwendungsbeispiel im Bereich der Information und Dokumentation ist der umfassende und systematische Einsatz von semantisch basierten Technologien beim Sender BBC, wie er in Lit. 21 ausführlich dokumentiert ist. Ein einheitlicher, „semantischer“ Ansatz für die Gestaltung der gesamten Web-Präsenz der BBC-Programme mit Angeboten für Endanwender und auch für Anwendungsentwickler führt dabei zu Gewinnen in den nachstehend benannten Bereichen:

- *Usability* – nutzerzentrierte Gestaltung von Web-Präsenzen.
- *User Experience* – Einsatz semantisch typisierter Prädikate und granular adressierbarer Entitäten, die solcherart in neuen Visualisierungsansätzen angeboten werden können.

- *User Journeys* – Das Angebot an Endnutzer, eigene Pfade durch das Angebot der BBC zu legen, denen nicht mehr ein traditioneller search&retrieval-Ansatz unterliegt, sondern die auf den Navigationstechniken der Linked Data aufbauen.
- *One page per thing* – Auflösung komplexer Seitenstrukturen in kleine, granulare Einheiten, die solcherart in der Präsenz der BBC dynamisch rekombinierbar und zugleich im Web der Linked Open Data direkt referenzierbar werden.
- *Our website is our API* – Die URIs der BBC-Präsenz sind Ankerpunkte für externe Anwendungen und den browserbasierten Zugriff. Die Präsenz wird damit nutzbar für Entwickler, die sie für die Gestaltung neuer Produkte verwenden können, wie das etwa in den Fällen URIPlay, TestTubeTelly, FanHubz or Channelography schon geschieht.
- *Loosely coupled development* – Unterschiedliche Entwicklerteams können lose verbunden zusammenarbeiten und ihre je spezifischen Interessengebiete bearbeiten.

Eine weitere Erfolgsgeschichte im Dokumentationsbereich ist der inzwischen fast flächendeckende Einsatz des SKOS-Standards des W3C (Lit. 22), mit dem inzwischen Normdaten aller Art von der Gemeinsam Normdatei der DDB (Lit. 20) bis zu den Arbeiten der „New Bibliographic Framework Initiative“ der Library of Congress, die vollständig auf RDF basieren, wie eine Anfrage unter <http://www.loc.gov/search/?q=resource+description+framework> ergibt.

Und auch ein großes Portal für die Repräsentation von Kulturgut im WWW wie die Europeana ist inzwischen in seiner RDF-basierten Version komplett über <http://europeana.ontotext.com/> in SPARQL adressierbar.

B 7.5 Das Ende des Semantic Web (?)

So hat es also den Anschein, als habe sich das Semantic Web nach einer Phase des Schwächelns in den mittleren Jahren 2000 und nicht zuletzt dank Linked Open Data auf breiter Front durchgesetzt. Und doch reißen die Diskussionen über ein baldiges Hinscheiden des Ansatzes nicht ab. Sie sind nun allerdings nicht mehr in der Kritik an der übertrieben KI-orientierten, logikdominierten Frühphase des Semantic Web verwurzelt, sondern haben einen anderen Ursprung, wie in der Invektive von Padley (Lit. 19). Die Annahme war dort, dass die von Google, Microsoft und Yahoo! ins Leben gerufene Initiative Schema.org das Semantic Web quasi links überholen und die syntaktische Komplexität und Differenziertheit von RDF zumindest für die Bedarfe der Suchmaschinenanbieter durch den wesentlich schlankeren Ansatz des Microdata-Formats und ein gemeinsames kontrolliertes Vokabular mit zumindest Ansätzen einer Zentralredaktion substituieren könnte. Ein Jahr nach Erscheinen der Polemik muss man allerdings feststellen, dass die Position von Schema.org sich zumindest aufzuweichen beginnt: Inzwischen wird dort RDFa mindestens auf einer Ebene mit Microdata gesehen, und auch der ursprünglich relativ exklusive Redaktionsansatz ist geöffnet worden.

Ebenfalls zu Spekulationen über ein Ende des Semantic Web Anlass gegeben hat dann im Frühjahr 2012 Googles Ankündigung, mit der Knowledge Graph Technologie (Lit. 10) eine „Revolution“ in der eigenen Suchmaschinentechnologie zu vollziehen. Knowledge Graph hat in der Tat die meisten Merkmale einer schlagkräftigen Semantic Web-Anwendung – inwieweit dahinter tatsächlich „semantische“ Technologien stehen, ist angesichts der Google-üblichen Intransparenz in solchen Dingen allerdings nicht zu beurteilen. Und doch erlaubt auch dieser Schachzug von Google zwei Interpretationen: Er kann als der Versuch gewertet werden, Semantic Web und Linked Data auf dem eigenen Feld (und womöglich mit den eigenen Mitteln) zu schlagen – genau so möglich ist allerdings die Interpretation, dass es sich hierbei um das implizite Eingeständnis des Suchmaschinen-giganten handelt, dass die eigene, rein statistisch basierte Methodik nicht mehr ausreicht und die RDF-basierten Ansätzen des Semantic Web schlicht ein größeres Potential haben.

Viel wesentlicher als diese letztlich derzeit nur spekulativ beantwortbare Frage ist allerdings eine andere, in den kommenden Jahren wahrscheinlich entscheidende und beispielsweise bei

Chandler et al. (Lit. 08) angedeutete Frage, diejenige nach dem spezifisch ökonomischen Mehrwert der semantischen Kontextualisierung! Ein reiches Forschungsfeld (gerade auch) für die Informationswissenschaft ... und für die weitere Entwicklung des Semantic Web als Linked Open Data.

Literatur

- 01 Antoniou, G.; Van Harmelen, F.: A Semantic Web Primer. (Cooperative Information Systems) (2nd ed.). MIT-Press, 2008 (<http://dl.acm.org/citation.cfm?id=1373341>)
- 02 Berners-Lee, T.: Linked Data – Design Issues, 2006 (<http://www.w3.org/DesignIssues/LinkedData.html>, retrieved December 16, 2012)
- 03 Berners-Lee, T.: Giant Global Graph. Decentralized Information Group (DIG) Breadcrumbs, 2007 (<http://dig.csail.mit.edu/breadcrumbs/node/215>)
- 04 Berners-Lee, T.: Information Management: a Proposal, 1989 (<http://www.w3.org/History/1989/proposal.html>, retrieved December 16, 2012)
- 05 Berners-Lee, T.: Artificial Intelligence and the Semantic Web. MIT, 2006 ([\(http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html#\(2\)\)](http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html#(2)))
- 06 Berners-Lee, T.: Linked Open Data, 2008 ([\(http://www.w3.org/2008/Talks/0617-lod-tbl/#\(1\)\)](http://www.w3.org/2008/Talks/0617-lod-tbl/#(1)))
- 07 Berners-Lee, T.; Hendler, J.; Lassila, O.: The Semantic Web. *Scientific American* 284 (5), 34-43, 2001 (doi:10.1038/scientificamerican0501-34)
- 08 Chandler; Vargo, S. L.: Contextualization and value-in-context: How context frames exchange. *Marketing Theory* 11 (1), 35-49, 2011 (doi:DOI: 10.1177/1470593110393713)
- 09 Gonzales, R.: Pharmaceutical Supply Chain Management Using Semantic Web Technology: Case Study, American Laboratory. American Laboratory, 2011 (www.americanlaboratory.com/914-Application-Notes/35621-Pharmaceutical-Supply-Chain-Management-Using-Semantic-Web-Technology-Case-Study/)
- 10 Google: Knowledge Graph – Alles über die Suche, 2012 (<http://www.google.com/insidesearch/features/search/knowledge.html>, retrieved December 16, 2012)
- 11 Gradmann, S.: Signal. Information. Zeichen. Zu den Bedingungen des Verstehens in semantischen Netzen. *Libreas*, 1-14, 2009
- 12 Gradmann, S.: Knowledge = Information in Context: on the Importance of Semantic Contextualisation in *Europeana*, 1-19, 2010 (<http://www.scribd.com/doc/32110457/Europeana-White-Paper-1>)
- 13 Harmelen, F. V.: Semantic Web Research anno 2006: main streams, popular fallacies, current status and future challenges, 2-8, 2006
- 14 Heath, T.; Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, 2011 (<http://linkeddatabook.com/editions/1.0/>)
- 15 Hendler, J.; Berners-Lee, T.: Artificial Intelligence – From the Semantic Web to social machines: A research challenge for AI on the World Wide Web 174 (2), 156-161, 2010 (dx.doi.org/10.1016/j.artint.2009.11.010)
- 16 Hitzler, P.; Krötzsch, M.; Rudolph, S.; Sure, Y.: *Semantic-Web-Grundlagen*. Heidelberg: Springer, 2008 (<http://www.semantic-web-grundlagen.de/>)
- 17 Nowack, Benjamin: The common, layered Semantic Web technology stack, n.d. (<http://smiy.wordpress.com/2011/01/10/the-common-layered-semantic-web-technology-stack/>)
- 18 O'Reilly, Tim, T.: Twitter / timoreilly: @joshu (twitter.com/timoreilly/status/1266750154)
- 19 Padley, R.: Triple bypass – What does the death of the semantic web mean for publishers? 2011 (<http://www.semantics.com/2011/09/triple-bypass-what-does-the-death-of-the-semantic-web-mean-for-publishers/>, retrieved December 16, 2012)
- 20 Pfeifer, B.: Gemeinsame Normdatei, 2012 (http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd_node.html, retrieved December 16, 2012)
- 21 Raymond, Y.; Scott, T.; Miller, L.; Betts, S.; McNamara: Case Study: Use of Semantic Web Technologies on the BBC Web Sites. W3C, 2012 (<http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/>)
- 22 World Wide Web Consortium: SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009 (<http://www.w3.org/TR/skos-reference/>)